Do Severity Measures Explain Differences in Length of Hospital Stay? The Case of Hip Fracture

Michael Shwartz, Lisa I. Iezzoni, Arlene S. Ash, and Yevgenia D. Mackiernan

Objective. To examine whether judgments about hospital length of stay (LOS) vary depending on the measure used to adjust for severity differences.
Data Sources/Study Setting. Data on admissions to 80 hospitals nationwide in the 1992 MedisGroups Comparative Database.
Study Design. For each of 14 severity measures, LOS was regressed on patient age/sex, DRG, and severity score. Regressions were performed on trimmed and untrimmed data. R-squared was used to evaluate model performance. For each severity measure for each hospital, we calculated the expected LOS and the z-score, a measure of the deviation of observed from expected LOS. We ranked hospitals by z-scores.
Data Extraction. All patients admitted for initial surgical repair of a hip fracture, defined by DRG, diagnosis, and procedure codes.
Principal Findings. The 5,664 patients had a mean (s.d.) LOS of 11.9 (8.9) days. Cross-validated R-squared values from the multivariable regressions (trimmed data) ranged from 0.041 (Comorbidity Index) to 0.165 (APR-DRGs). Using untrimmed data, observed average LOS for hospitals ranged from 7.6 to 23.9 days. The 14 severity measures showed excellent agreement in ranking hospitals based on z-scores. No severity measure explained the differences between hospitals with the shortest and longest LOS.
Conclusions. Hospitals differed widely in their mean LOS for hip fracture patients, and severity adjustment did little to explain these differences.
Key Words. Severity, length of stay, hospital efficiency, hip fracture

Whether or not severity of illness is important when considering resource consumption of hospitalized patients has been a persistent policy and research question since Medicare adopted diagnosis-related groups (DRGs) for prospective hospital payment in 1983 (Vladeck 1984; Jencks and Dobson 1987; Edwards et al. 1994). For hospital payment, the major issue is whether patient severity differs systematically at hospitals in ways not captured by DRGs, and whether these differences explain variation in resource use.

This question remains despite the development of various severity measurement approaches in the last decade. Many articles about these measures, written primarily by their developers, have appeared in the clinical and health services research literature (see further on). However, little has been published by independent investigators comparing predictions of resource use across severity measures (Thomas, Ashcraft, and Zimmerman 1986; Thomas and Ashcraft 1991; MacKenzie, Willan, Lichter, et al. 1991). This study applied 14 measures to a single data set, to see whether judgments about length of stay (LOS) are sensitive to the severity measure used to adjust for severity differences. We examined LOS of patients admitted to 80 hospitals for hip fracture repair.

BACKGROUND

Questions linking severity to fairness of hospital reimbursement arose even before Medicare adopted DRG-based payment. In his *Report to Congress* proposing prospective payment, Secretary of Health and Human Services Richard S. Schweiker (1982) acknowledged that severity differences within DRGs could raise problems if certain hospitals had more severely ill cases than others and if severity is positively correlated with hospitalization costs. Concerns focused particularly on hospitals generally viewed as attracting sicker patients, such as tertiary teaching centers and public institutions caring for the indigent.[1]

In response, the 1986 Omnibus Budget Reconciliation Act (P.L. 99–509) required that by 1988 the Administration propose a method "to account for

---

This research was supported by the Agency for Health Care Policy and Research, under Grant No. RO1 HS06742–03. The views expressed are solely those of the authors.

Address all correspondence and requests for reprints to Lisa I. Iezzoni, M.D., M.Sc., Associate Professor of Medicine, Division of General Medicine and Primary Care, Department of Medicine, Beth Israel Hospital, Libby Building Room 326, 330 Brookline Avenue, Boston, MA 02215. Michael Shwartz, Ph.D. is Associate Professor, Health Care Management Program and Operations Management Program, School of Management, Boston University. Arlene S. Ash, Ph.D. is Associate Research Professor in the Health Care Research Unit, Section of General Internal Medicine, Evans Memorial Department of Clinical Research and Medicine, Boston University Medical Center. Dr. Iezzoni and Yevgenia D. Mackiernan, Programmer, are affiliated with the Division of General Medicine and Primary Care, Department of Medicine, Harvard Medical School, Beth Israel Hospital, the Charles A. Dana Research Institute, and the Harvard-Thorndike Laboratory. This article, submitted to *Health Services Research* on May 4, 1995, was revised and accepted for publication on February 12, 1996.

variations in severity of illness and case complexity which are not adequately accounted for by the current classification and payment system." How to do so was unclear. After studying existing severity measures, the Health Care Financing Administration (HCFA) Office of Research and Demonstrations (1987) reported, "The Department does not believe that any system to measure clinical severity of illness is currently an administratively feasible major improvement to or substitute for DRGs." HCFA funded a comparative study of different severity measures that yielded inconclusive results (MacKenzie, Willan, Greenaway-Coates, et al. 1991). HCFA also supported researchers at Yale University, including several of the original DRG developers, to create a "refined" version (R-DRGs) that was more sensitive to severity differences (Fetter, Freeman, Park, et al. 1989; Freeman, Fetter, Park, et al. 1995).

The issue was reopened in HCFA's Proposed Rules for the Medicare program published in the May 26, 1993, *Federal Register* (p. 30230):

> For several years, we have been analyzing major refinements to the DRG classification system to compensate hospitals more equitably for treating severely ill Medicare patients. These refinements, generally referred to as severity of illness adjustments, would be based on using certain complications and comorbidities (CCs) to create DRGs specifically for very ill patients who consume far more resources than do other patients classified to the same DRG in the current system. This approach has been taken by various other groups in refining the DRG system, most notably the research done for Yale and the changes incorporated by the State of New York into its all-payer DRG system.

In mid-1994, HCFA released a description of its new severity adjustment approach, soliciting public comments.[2] The proposed approach retains the DRG structure (Edwards et al. 1994). HCFA explicitly rejected using clinical data-based severity measures because of cost, and it did not revisit the DRGs' fundamental conceptual model of severity (Iezzoni and Moskowitz 1986). Therefore, although the federal government may move ahead with a revised version of the DRGs, this approach probably will not silence critics concerned about systematic severity differences across hospitals.

METHODS

Severity Methodologies. We considered 14 severity methodologies (Table 1). These measures are representative of approaches used currently in severity-adjusting outcomes data for hospital performance reports, individual hospital activities (e.g., internal monitoring, negotiating contracts with managed care organizations), and clinical and health services patient outcomes research

Table 1: Description of Severity Methods

Severity Method	*Source*	*Data Used and Definition of Severity**	*Classification Approach and Derivation†*
Discharge Abstract–Based Methods: Resource-Driven Severity Definition			
Acuity Index Method (AIM)[1]	Iaméter	Discharge abstract; length of hospital stay within DRGs	Scores 1, 2, 3, 4, or 5 within DRG;‡ empirical modeling
All Patient Refined Diagnosis Related Groups (APR-DRGs)[2]	3M Health Information Systems	Discharge abstract; total hospital charges	Four severity classes (A, B, C, D) within adjacent DRGs;§ empirical modeling with clinical guidance
Disease Staging Relative Resource Scale	SysteMetrics/MEDSTAT Group	Discharge abstract; relative total hospital charges	Relative weight with 100 as an average
Patient Management Categories (PMCs) Relative Intensity Score (RIS)[3]	Pittsburgh Research Institute	Discharge abstract; relative resource consumption	Weight compared to an average of 1.0; empirical modeling with clinical guidance
Refined Diagnosis-Related Groups (R-DRGs)[4]	Yale University refinement of DRGs provided by Yale Project Director Karen Schneider, now with Health Systems Consultants (New Haven, CT)	Discharge abstract; length of hospital stay, total hospital charges	Three severity classes (B, C, D) within adjacent medical DRGs;[d] "early" deaths grouped in lowest severity class; empirical modeling with clinical guidance
Discharge Abstract–Based Methods: Clinical Severity Definition			
Disease Staging mortality probability	SysteMetrics/MEDSTAT Group	Discharge abstract; probability of in-hospital death	Probability ranging from 0 to 1; empirical modeling
Disease Staging stage[5]	SysteMetrics/MEDSTAT Group	Discharge abstract; risk of death or functional impairment	Three stages (1.0, 2.0, and 3.0) with substages within each stage; clinical judgment

Disease Staging comorbidity count[6]	SysteMetrics/MEDSTAT Group	Discharge abstract; number of comorbid conditions within each of the three major stages	Three separate variables: the integer associated with each of the three stages (1, 2, 3); clinical judgment
Patient Management Categories (PMCs) Severity Score[7]	Pittsburgh Research Institute	Discharge abstract; in-hospital morbidity and mortality	Score of 1, 2, 3, 4, 5, 6, or 7; empirical modeling
Comorbidity Index	Developed by Charlson et al.;[8] coded version patterned after Deyo et al.[9]	Discharge abstract; risk of death within one year of medical hospitalization	Integer from additive scale representing number and severity of comorbidities; clinical judgment with empirical guidance
Body Systems Count[10]	H.C.I.A., Inc.	Discharge abstract; number of organ systems involved with disease	Integer count; clinical judgment
Clinical Data-Based Methods: Clinical Severity Definition			
MedisGroups (Atlas MQ)	MediQual Systems, Inc.	Clinical data	
Original version[11]		Clinical instability indicated by in-hospital death; score independent of diagnosis	Admission score 0, 1, 2, 3, or 4; clinical judgment
Empirical version[12]		In-hospital death; score calculated within 64 disease groups	Probability ranging from 0 to 1; empirical modeling
Physiology Score 1	Patterned after Acute Physiology Score, Acute Physiology and Chronic Health Evaluation II (APACHE II)[13]	Clinical data; in-hospital mortality for patients in intensive care unit	Integer score starting with 0; APACHE II's Acute Physiology Score ranges from 0 to 60; clinical judgment with empirical guidance
Physiology Score 2	Patterned after Acute Physiology Score, Acute Physiology and Chronic Health Evaluation III (APACHE III)[14]	Clinical data; in-hospital mortality for patients in intensive care unit	Integer score starting with 0; APACHE III's Acute Physiology Score ranges from 0 to 252; empirical modeling with clinical guidance

Continued

Table 1: Continued

[*] "Discharge abstract" indicates standard hospital discharge data elements, such as basic demographics, diagnosis, and procedure codes. "Clinical data" indicates clinical information (e.g., vital signs, test results) abstracted from the medical record.

[†] "Derivation" indicates the principal method used to create the severity scoring method. "Clinical judgments" reflects primarily use of expert physician guidance. "Empirical modeling" indicates primarily use of statistical techniques.

[‡] Diagnosis-related group.

[§] "Adjacent DRGs" are formed by grouping individual DRGs previously split by complications and comorbidities.

[1] Iezzoni 1994; Thomas and Ashcraft 1991.

[2] All Patient Refined Diagnosis Related Groups, 3M Health Information Systems 1993.

[3] Young 1984; Young, Swinkola, and Zorn 1982.

[4] Freeman, Fetter, Park, et al. 1991; Health Systems Management Group, School of Organization and Management, Yale University 1989.

[5] Gonnella, Hornbrook, and Louis 1984; Markson et al. 1991; Gonnella et al. 1990.

[6] Naessens et al. 1992.

[7] Young, Kohler, and Kowalski 1994.

[8] Charlson et al. 1987.

[9] Deyo, Cherkin, and Ciol 1992.

[10] Mendenhall 1984.

[11] Brewster, Karlin, Hyde et al. 1985; Iezzoni and Moskowitz 1988; Blumberg 1991.

[12] Steen, Brewster, Bradbury et al. 1993.

[13] Knaus, Wagner, Draper et al. 1985; Knaus et al. 1986.

[14] Knaus, Wagner, Draper et al. 1991; Knaus, Wagner, and Lynn 1991; Knaus et al. 1993.

(Iezzoni, Shwartz, and Restuccia 1991; Iezzoni and Greenberg 1994). Each approach defines "severity," reflecting how it was derived and calibrated (Table 1). Measures can be grouped into those focusing on clinical outcomes (typically mortality) and those targeting resource consumption. We included mortality-based measures because they are more likely to represent clinicians' thinking about "severity" than do resource-based approaches.

Ten measures (all except the two MedisGroups and two Physiology Scores) rate severity using standard data elements from hospital discharge abstracts (U.S. Department of Health, Education and Welfare, National Committee on Vital and Health Statistics 1980; Anderson, Steinberg, Whittle, et al. 1990; Connell, Diehr, and Hart 1987; Roos et al. 1989), such as age, sex, and diagnoses and procedures coded using the *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM). MedisGroups and the Physiology Scores assess severity using clinical data (e.g., on vital signs, laboratory findings) abstracted from medical records. Severity measures assign either numerical scores or values on a continuous scale (see Table 1).

Database. To assign severity scores, computerized algorithms were applied to a data file extracted from the 1992 MedisGroups® Comparative Database (Iezzoni, Shwartz, Ash, et al. 1995). Briefly, this database contains the clinical information collected on hospitalized patients during medical record reviews using the MedisGroups severity measure (Brewster, Karlin, Hyde, et al. 1985; Iezzoni and Moskowitz 1988; Steen, Brewster, Bradbury, et al. 1993). The 1992 database contains all calendar year 1991 discharges from 108 acute care hospitals thought to have good data and representing a range of characteristics.

To ensure adequate numbers for hospital-level analyses, we selected the 80 hospitals with the most hip fracture patients, eliminating facilities with fewer than 29 patients (a total of 421 patients). Fifty-five percent of the hospitals were in Pennsylvania; 51.2 percent had more than 300 beds; 93.7 percent were private, nonprofit; and 48.8 percent had approved residency training programs (Iezzoni, Shwartz, Ash, et al. 1995).

Original and empirical MedisGroups scores were provided by MediQual Systems in the file given to us, while scores for other measures had to be assigned. The MedisGroups Comparative Database includes the values of the key clinical findings (KCFs) abstracted from medical records during admission MedisGroups reviews. This KCF information was used to create physiology scores patterned after APACHE® II and III. Physiologic findings were given weights specified by APACHE II and III; for example, for APACHE III, a pulse of 145 beats/minute generates 13 points (Knaus, Wagner, Draper, et al. 1991). As with APACHE II and III, these weights were summed to produce the score. We could not replicate actual Acute Physiology Scores of APACHE II and III because complete values for the required physiologic variables were not available: MedisGroups truncates data collection in a broadly defined normal range (Iezzoni, Hotchkin, Ash, et al. 1993).

The MedisGroups database also contains standard discharge abstract information assigned by the hospitals, including ICD-9-CM codes for up to 20 diagnoses and 50 procedures. For the ten discharge abstract–based severity measures, we assigned only the code-based version of the Charlson comorbidity index (Charlson et al. 1987), using an approach adapted from ICD-9-CM diagnosis codes developed by Deyo, Cherkin, and Ciol (1992). Other severity scoring was performed by the vendors (Table 1), using computer files containing the necessary discharge abstract data elements extracted by us from the MedisGroups database. We used version 9.0 of the Medicare DRGs, as assigned by MediQual Systems. We merged scores of the different measures into a single analytic file with complete success.

Study Sample and Outcome Measure. We studied LOS of patients hospitalized for surgical treatment of hip fracture by first identifying cases with a principal ICD-9-CM diagnosis code beginning with 820 (fracture of neck of the femur). We then eliminated cases admitted only for removal of an internal fixation device or prosthesis, not for initial surgical repair. To include only patients treated surgically for hip fracture, we retained cases in DRGs 209 (major joint procedures of lower extremity), 210 (hip and femur procedures except major joint, with complication or comorbidity [CC]), and 211 (hip and femur procedures except major joint, without CC; see Table 2).

ANALYTIC METHODS

We examined two major categories of questions:

1. To what extent did different severity measures explain differences in LOS *at the individual patient level*? and
2. To what extent did different severity measures explain differences in LOS *at the hospital level*?

Predictive Models. Both analyses relied on multivariate regression models, with severity score entered as one or more independent variables along with 15 dummy variables for patient sex interacted with eight age categories (18–44, 45–54, 55–64, 65–69, 70–74, 75–79, 80–84, and 85 years of age or older) and two dummy variables for DRG.

Table 2: Selected Characteristics of Patient Sample by DRG

Characteristic	*DRG* 209	210	211	*All Cases*
Number of cases	1,981	2,839	844	5,664
Percent female	78.8	76.0	76.1	77.0
Percent died in-hospital	3.5	3.7	0.1	3.1
Percent with Medicare as primary payer	90.7	89.1	76.7	87.8
Mean (s.d.) age (in years)	79.8 (9.0)	79.9 (11.4)	75.0 (15.2)	79.1 (11.5)
Mean (s.d.) length of stay (in days)*	12.2 (9.4)	12.5 (9.5)	8.9 (3.9)	11.9 (8.9)
Mean (s.d.) total charges (in dollars)	17,463 (10,919)	14,979 (10,364)	10,419 (4,230)	15,164 (10,168)
Mean (s.d.) diagnosis codes	5.6 (3.0)	6.3 (2.8)	3.2 (1.7)	5.6 (2.9)

*Untrimmed data.

Separate models were calculated for each severity measure, with scores entered as either continuous or categorical variables (Table 1). For empirical MedisGroups and the mortality version of Disease Staging, both of which produce a "predicted probability of death" (p) as the severity score, we included both p and p^2 in the model. This allowed models to reflect the possibility that both the least severe (low predicted probability of death) and most severe (high probability) patients have relatively short lengths of stay.

Three measures (AIM, APR-DRGs, and R-DRGs) consider DRGs (or some variant of DRGs) in rating patients. For these measures, we analyzed additional models with interaction terms for severity crossed with DRG. The interaction variable improved model fit only for AIM; we therefore present AIM results from this interacted model. For the two other measures, we report only results from entering DRG and severity, as separate independent variables, and not their interaction terms.

Transformations and Outliers. We fit models to predict both LOS and the natural logarithm of LOS. When using either outcome, model fit (i.e., the R-squared value) was evaluated on the scale of days, rather than log(days). Duan's smearing estimator (Duan 1983) was used to re-transform predicted LOS from the log scale. Since R-squared values were as high or higher using LOS rather than log (LOS) as the dependent variable, we report results for using LOS only.

We performed analyses including all data (untrimmed) and after removing outliers (trimmed). Two methods were used to identify outliers: (1) the

HCFA approach, trimming all cases more than three standard deviations from the mean on a log scale; and (2) a robust method for outlier identification proposed by Hoaglin and Iglewicz (1987).

Individual Patient-Level Analyses. We used R-squared as an overall measure of the ability of each model to predict individual patient outcomes. To avoid overly optimistic assessments of model performance due to over-fitting, we used split-sample cross-validation. First, the data were randomly split in half and a model estimated on each subsample. Then, a "validated" R-squared was computed in each half by comparing actual LOS to predictions made using the model fit to the other half. The cross-validated R-squared was calculated as the average of these two R-squared values.

To illustrate each measure's ability to discriminate low from high LOS cases, we ranked patients from lowest to highest predicted LOS and divided the data into deciles. We report the average LOS for the lowest and highest two deciles.

Hospital-Level Analyses. Using each severity measure in each of the 80 hospitals, we estimated the expected LOS and its standard error. Expected LOS was calculated as the average, across patients within a facility, of individual predicted LOS from the multivariate regression model for the particular severity measure. The standard error was computed as the square root of the sum of the squares of the standard errors from the regression equations for each patient. A z-score was calculated for each hospital as $z =$ (observed average LOS $-$ expected LOS)/(standard error of LOS). We then ranked hospitals from lowest (shorter LOS than expected) to highest (longer LOS than expected) based on these z-scores. When considering hospital LOS unadjusted for age, sex, or severity, expected LOS was set equal to the mean LOS across all 5,664 patients.

We examined three measures of hospital performance looking at whether each hospital

1. ranked among the worst 10 percent (the 8 hospitals with the highest z-scores);
2. ranked among the best 50 percent (the 40 hospitals with the lowest z-scores); and
3. was a statistical outlier (z-score > 2 or < -2, indicating significantly longer or shorter LOS observed than expected).

For each measure, a severity measure either "flagged" a hospital (e.g., identified the hospital as among the worst 10 percent) or it did not. For each pair of severity measures, we examined the number of hospitals classified similarly by both measures (i.e., either flagged or not) and calculated a kappa statistic. Kappa measures the extent to which two severity measures agree on flagging hospitals more than expected by chance. Kappa values below 0.4 generally indicate poor to fair agreement, while values greater than 0.7 usually mean excellent agreement (Landis and Koch 1977). Findings from the unadjusted model and the model using only age-sex and DRG were included in these pairwise comparisons.

To illustrate how severity measures might "explain" the observed differences among hospitals in average LOS, we examined the three hospitals with the lowest LOS, the two median hospitals in terms of LOS, and the three hospitals with the highest LOS. For each hospital, we compared actual to predicted LOS for each of the severity measures.

To summarize the ability of severity to explain differences in LOS among hospitals, we examined the change in R-squared when each severity measure was added to a model that included the age-sex category and DRG plus dummy variables for each hospital. We also examined the correlation between actual and predicted hospital LOS for each measure.

RESULTS

The analytic file contained 5,664 patients from 80 hospitals. Patients ranged from 18 to 104 years of age, with a mean (standard deviation) age of 79.1 (11.5) years; 77.0 percent of the patients were female (Table 2). Almost 13 percent of the patients were admitted from nursing homes or skilled nursing facilities, and 3.6 percent were transferred from another acute care hospital. Mean (s.d.) LOS was 11.9 (8.9) days. Only 5 percent of patients stayed fewer than six days; another 5 percent stayed more than 24 days. For the discharge abstract–based measures, ample numbers of diagnosis codes were generally present for rating severity. The average case had 5.6 (s.d. = 2.9) diagnosis codes listed. Only 4.1 percent of patients had just one discharge diagnosis; 45.7 percent had more than five diagnoses, and 10.0 percent had ten or more.

The 80 hospitals treated 29 to 182 cases, with a mean of 70.8 (s.d. = 34.6) cases and a median of 64 cases. Average hospital LOS varied widely, from 7.6 days to 23.9 days. For major joint replacements (DRG 209), the mean LOS per hospital ranged from 7.8 days to 33.4 days; for other hip fracture surgery (DRGs 210 and 211), the mean LOS per hospital was 7.5 to 29.1 days.

Performance in Predicting LOS. Using HCFA's approach to identify outliers, we removed 57 patients from the data set, reducing average LOS from

11.9 days to 11.5 days for the trimmed data. The Hoaglin approach eliminated many more cases (414), shortening average LOS to 10.3 days. R-squared was higher using Hoaglin trimming for only one of the 14 severity measures; for eight measures, R-squared was almost 50 percent higher using the HCFA trim points. Results reported here use HCFA's definition of outliers.

The 14 severity measures varied in their statistical performance (Table 3). On trimmed data, APR-DRGs had the highest R-squared (0.171), followed by AIM (0.147, interacting severity score with DRG). The Comorbidity Index had the lowest R-squared (0.051), only slightly higher than the age-sex and DRG model (0.045). Cross-validated R-squared values were lower, but generally by a small amount. For all measures, models performed substantially better on trimmed data than on untrimmed.

For all measures, cases in the lowest decile of predicted LOS had mean LOSs between 8.3 and 9.1 days, while cases in the highest decile had mean LOSs from 13.1 to 17.3 days (for APR-DRGs). Only for the Comorbidity Index and the age-sex-DRG were the LOSs in the highest decile less than 50 percent greater than LOSs in the lowest decile.

Relative Hospital Performance. For each pair of measures, we examined the number of hospitals classified as among the worst 8 and the best 40. All pairs of measures agreed on at least 7 of the worst 8 hospitals (kappa 0.86); they also agreed on at least 36 of the hospitals in the 40 best (kappa 0.80). Rankings based on observed LOS corresponded well with severity-adjusted rankings. The average kappa resulting from comparing rankings based on observed LOS with rankings based on severity-adjusted LOS across the different severity measures was 0.92 for flagging the 8 worst hospitals and 0.90 for flagging the 40 best.

Mean LOS differed by hospital (ANOVA, $p < .001$), with the shortest mean LOS equaling 7.6 days and the longest 17.7 days. Predicted LOS did not capture differences in LOS among hospitals at the extremes (Table 4). The shortest LOS predicted by any of the measures for the three shortest LOS hospitals was 10.8 days; the longest LOS predicted for the three longest-stay hospitals was 13.0 days. The largest difference between highest and lowest LOS predictions was 3.3 days (Body Systems); for many measures the range of predicted LOS was under two days. Across the severity measures, predicted LOSs for the three hospitals with the shortest stays were almost indistinguishable from the predicted LOSs for the three hospitals with the longest stays.

The R-squared in a model with DRG, age-sex, and dummy variables distinguishing hospitals was 14.6 percent, of which 10.0 percent was due to the hospital dummy variables. We added each severity measure to this model

Table 3: *R*-Squared Values and Actual Average Length of Stay for Highest and Lowest Deciles Based on Predicted Lengths of Stay by Severity Method

	R-Squared Values × 100			*Average Length of Stay Decile*†			
Severity Method	*All Cases*	*Trimmed**	*Trimmed, Cross-Validated*	*1*	*2*	*9*	*10*
				average LOS in days for patients in decile			
No adjustment				11.5	11.5	11.5	11.5
Age-sex, DRG	2.4	4.5	3.7	8.7	10.2	12.1	13.1
AIM(1)‡	7.0	12.0	–§	8.8	9.7	13.3	15.8
AIM(2)	9.1	14.7	11.7	8.6	9.5	13.4	15.8
APR-DRG(1)‡	11.0	17.4	16.5	9.0	9.3	13.1	17.3
APR-DRG(2)	10.9	17.1	–§	9.0	9.8	13.2	17.3
DS Relative Resource Scale	6.9	11.6	10.9	8.3	10.1	13.0	15.8
PMC Resource Intensity Score	6.2	11.0	10.3	8.7	9.8	13.2	15.6
R-DRG(1)‡	7.5	11.8	11.3	9.1	9.5	13.0	15.6
R-DRG(2)	7.7	11.8	–	9.0	9.5	13.0	15.6
DS Mortality Probability	4.6	8.8	8.0	8.7	10.1	13.1	15.4
DS Stage	2.9	5.0	4.2	8.7	10.2	12.2	13.4
DS Comorbidity Count	7.4	12.4	11.7	8.7	9.8	13.6	15.6
PMC Severity Score	5.5	9.7	8.7	8.8	9.7	12.5	15.5
Comorbidity Index	2.6	5.1	4.1	8.7	10.1	12.3	13.5
Body Systems Count	7.7	11.6	10.7	8.6	9.7	13.4	15.5
MedisGroups original	3.5	6.7	5.6	8.7	10.1	12.7	13.9
MedisGroups empirical	2.9	5.8	4.8	8.8	10.0	12.9	13.8
Physiology Score 1	3.4	6.7	5.9	8.6	10.2	12.8	14.0
Physiology Score 2	3.2	6.0	5.1	8.8	10.0	12.4	13.8

Note. AIM = Acuity Index Method; APR-DRG = All Patient Refined DRGs; DS = Disease Staging; PMC = Patient Management Category; R-DRG = Refined DRGs.

*Cases outside ±3 standard deviations on log scale eliminated.

†Trimmed data.

‡(1) Model in which severity score is not interacted with DRG; (2) model in which severity score is interacted with DRG.

§Cross-validated R^2 were calculated on the best of the two models only.

and re-examined the contribution to *R*-squared due to hospitals after the other variables had been entered. For most measures, *R*-squared attributed to hospitals remained essentially unchanged following case-mix adjustment.

Table 4: Expected Lengths of Stay by Severity System for Hospitals with Lowest, Median, and Highest Actual Lengths of Stay (LOS)

	*Hospitals grouped by actual LOS**								
	Lowest LOS			*Median LOS*		*Highest LOS*			
Adjusted for Indicated Variables or Severity Scores	*A* (n = *31*)	*B* (n = *47*)	*C* (n = *64*)	*D* (n = *61*)	*E* (n = *90*)	*F* (n = *33*)	*G* (n = *77*)	*H* (n = *30*)	*Range of LOS*
Average LOS (trimmed)	7.6	8.4	8.6	11.2	11.3	16.2	16.2	17.7	10.1
Expected LOS									
No adjustment	11.5	11.5	11.5	11.5	11.5	11.5	11.5	11.5	
Age-sex, DRG	11.7	11.0	11.3	11.3	11.8	11.3	11.7	11.7	1.5
AIM†	11.9	11.4	12.7	11.6	11.8	11.3	12.7	11.0	2.8
APR-DRG	11.5	11.2	11.0	11.9	11.5	12.0	12.5	11.2	2.4
DS Relative Resource Scale	12.2	11.5	11.2	11.6	11.5	11.6	13.0	11.2	2.6
PMC Resource Intensity Score	12.3	11.0	10.8	11.3	11.7	11.6	12.4	11.5	2.6
R-DRG	11.8	10.9	11.1	11.7	11.6	11.3	12.2	11.3	2.2
DS Mortality Probability	11.4	11.1	11.0	11.4	11.5	11.2	11.9	11.4	1.5
DS Stage	11.6	11.1	11.2	11.4	11.7	11.2	11.8	11.7	1.8
DS Comorbidity Count	12.2	11.2	10.9	11.7	11.2	11.6	13.0	11.3	2.8
PMC Severity Score	12.1	11.0	11.0	11.3	11.8	11.6	12.3	11.1	1.9
Comorbidity Index	11.6	11.1	11.2	11.3	11.7	11.3	11.9	11.6	1.5
Body Systems Count	12.0	11.2	10.9	11.3	11.3	11.4	12.9	11.3	3.3
MedisGroups original	11.7	10.8	11.1	11.4	11.7	11.6	11.7	11.8	1.6
MedisGroups empirical	11.7	11.2	11.1	11.3	11.7	11.3	11.7	11.5	1.7
Physiology Score 1	11.8	11.1	11.2	11.4	11.7	11.5	11.6	11.5	1.5
Physiology Score 2	11.6	11.1	11.3	11.3	11.7	11.3	11.6	11.7	1.6

Note: AIM = Acuity Index Method; APR-DRG = All Patient–Refined DRGs; DS = Disease Staging; PMC = Patient Management Category; R-DRG = refined DRGs.

* Actual LOSs used here are calculated on trimmed data.

† In the AIM model severity score is interacted with DRG; in other models, severity score is not interacted with DRG.

The largest change was when using APR-DRGs, but even there it fell only 0.5 percent, to 9.5 percent. Further, correlation between actual and predicted LOS was small for all measures. For APR-DRGs, the measure with the highest R-squared at the individual patient level, the correlation at the facility level was 0.24. Thus, no severity measure accounted for much variation in hospital LOS.

DISCUSSION

Severity measures differed in their ability to predict LOS at the individual patient level, with validated R-squared values ranging from 4.9 percent to 17 percent. More important from a policy perspective, however, was that no measure indicated large differences in expected hospital LOS due to patient severity mix, and none of the 14 severity measures provided much explanation of the wide differences observed across hospitals in average LOS.

The 14 severity measures substantially agreed in ranking the "best" and "worst" hospitals by observed-to-expected LOS. However, this agreement resulted from their tendency to view all hospitals as having a similar severity mix; identifying the set of "best" and "worst" facilities based on *observed* (unadjusted) LOS agreed as well with severity-adjusted rankings as the severity-adjusted rankings agreed with each other. This further confirms that, at the hospital level, severity adjustment added little to our understanding of which hospitals had unexpectedly high (or low) mean LOSs. Observed differences in average LOS reflect both "real" differences across hospitals and random variation; however, these data clearly exhibit large differences in LOS that are not explained by severity adjustment.

Results reported here pertain to one condition only–initial surgical treatment for hip fracture. We chose this condition for several reasons. Making the diagnosis raises few questions, ICD-9-CM hip fracture codes are fairly clear, and the treatment is generally straightforward. Differences across treatments are reasonably proxied by DRGs (209 for hip replacement and 210/211 for internal fixation). Compared to other inpatient conditions, such as cardiac disease, there are relatively few applicable expensive diagnostic procedures that can be performed at physicians' discretion, thus increasing LOS variation because of provider practice patterns rather than patient need. Although hip fracture patients generally do not experience the serious physiologic derangements of acute myocardial infarction, for example, comorbid disease may be a critical factor for persons with hip fractures. Comorbid illness may be captured by ICD-9-CM diagnosis codes, thereby giving code-based severity measures (that do not have access to physiologic values) an opportunity to perform. However, LOS for a hip fracture patient is likely to be influenced by prefracture functional status–an important attribute not measured by any of the 14 severity measures. Thus, our results may not generalize to other conditions.

The wide variations in lengths of hospital stays remained largely unexplained. We examined the possibility that the availability of rehabilitation services for non-acute care placement may be associated with shorter LOSs for certain hospitals. However, many hospitals never used the rehabilitation disposition code (30 of the 80 hospitals indicated no patients discharged to rehabilitation, probably reflecting coding practices rather than reality). For those hospitals using the rehabilitation disposition code, there was little relationship between average LOS and percent discharged to rehabilitation. For example, in the hospital with the shortest average LOS, 58 percent of patients were discharged to rehabilitation units; in the hospital with the longest LOS, 52 percent were discharged to rehabilitation units.

Thomas and Ashcraft (1991) have been among the few to report R-squared values using a variety of severity measurements to predict hospital costs. Their sample included patients from 11 groups of adjacent DRGs, ranging from simple pneumonia and pleurisy (DRGs 89–91) to coronary artery bypass surgery (DRGs 106 and 107) and diabetes (DRGs 294 and 295). Weighted R-squared values across conditions (the only values they report) were generally similar to those we report, ranging from 0.070 to 0.203.

Some severity measures were explicitly aimed to predict LOS or some other measure of resource consumption, while other measures were to predict mortality or a more clinical perspective on severity. No single definition is absolutely correct: the term *severity* alone has little intrinsic meaning (Iezzoni 1994). We included mortality-based measures because they generally reflect how clinicians think about severity better than resource-based measures do. Given that HCFA is leaning toward a resource-driven, DRG-based severity measure (Edwards et al. 1994), it is unlikely that HCFA's new severity adjustment will satisfy clinicians concerned about severity differences among patients.

This study had several limitations, primarily pertaining to the data. The database contains information only from self-selected purchasers of MedisGroups or from hospitals in states with MedisGroups data collection mandates. Data reliability could not be independently verified, and the clinical information contained in the data set was specifically gathered for MedisGroups scoring. Most importantly, it did not contain functional status information. Despite this, the MedisGroups Comparative Database is the only available database that includes reasonably detailed clinical data on all cases, regardless of insurer, from a range of hospitals across the country. Numerous coding slots were available for the most important discharge abstract data, ICD-9-CM diagnosis codes.

We included the physiology scores not specifically to examine APACHE itself, but because of the increasing interest in creating "minimum clinical data sets" containing a small number of well-selected, physiologic variables. APACHE weights represent one way to use these minimum physiologic variables, but there are certainly other approaches. Physiology Score 1 considered 12 variables, while Physiology Score 2 involved 17 items. It is important to note that the MedisGroups data abstraction protocol collects a generic set of over 200 potential "key clinical findings" regardless of patient diagnosis. The performance of the physiology scores is often comparable to that of empirical MedisGroups, but neither of these two clinical data–based measures explained LOS. A new version of MedisGroups is being empirically derived explicitly to predict LOS, and an LOS-based version of APACHE III is currently available (Knaus et al. 1993).

Our results suggest that severity adjustment is of little value in identifying the "best" and "worst" hospitals in terms of average LOS for hip fracture patients. Rankings based on observed LOS with no severity adjustment convey similar information. Nonetheless, severity adjustment may be valuable for utilization review and continuous improvement efforts within individual hospitals, for it is able to identify groups of patients with lengths of stay that differ by 50 percent to 100 percent.

NOTES

1. Certain technical provisions of Medicare's prospective payment formula (e.g., outlier, disproportionate share, and indirect medical education payments) served partially as proxy solutions for specific areas where severity could be important.
2. HCFA's proposed DRG refinement is described in a 1994 unpublished paper, "Refinement of the Medicare Diagnosis-Related Groups to Incorporate a Measure of Severity." The computer software required to calculate these new DRGs was not available to us.

REFERENCES

All Patient Refined Diagnosis Related Groups. Definition Manual. 1993. Wallingford CT: 3M Health Information Systems.

Anderson, G., E. P. Steinberg, J. Whittle, N. R. Powe, S. Antebi, and R. Herbert. 1990. "Development of Clinical and Economic Prognoses from Medicare Claims Data." *Journal of the American Medical Association* 263: 967–72.

Averill, R. F., T. E. McGuire, B. E. Manning, D. A. Fowler, S. D. Horn, P. S. Dickson, M. J. Coye, D. L. Knowlton, and J. A. Bender. 1992. "A Study of the Relationship Between Severity of Illness and Hospital Cost in New Jersey Hospitals." *Health Services Research* 27 (5): 587–606.

Blumberg, M. S. 1991. "Biased Estimates of Expected Acute Myocardial Infarction Mortality Using MedisGroups Admission Severity Groups." *Journal of the American Medical Association* 265 (22): 2965–70.

Bradbury, R. C., J. H. Golec, and F. E. Stearns. 1991. "Comparing Hospital Length of Stay in Independent Practice Association HMOs and Traditional Insurance Programs." *Inquiry* 28 (1): 87–93.

Brewster, A. C., B. G. Karlin, L. A. Hyde, C. M. Jacobs, R. C. Bradbury, and Y. M. Chae. 1985. "MEDISGRPS: A Clinically Based Approach to Classifying Hospital Patients at Admission." *Inquiry* 12 (4): 377–87.

Charlson, M. E., P. Pompei, K. L. Ales, and C. R. MacKenzie. 1987. "A New Method of Classifying Prognostic Comorbidity in Longitudinal Studies: Development and Validation." *Journal of Chronic Diseases* 40 (5): 373–83.

Connell, F. A., P. Diehr, and L. G. Hart. 1987. "The Use of Large Data Bases in Health Care Studies." *Annual Review of Public Health* 8: 51–74.

Deyo, R. A., D. C. Cherkin, and M. A. Ciol. 1992. "Adapting a Clinical Comorbidity Index for Use with ICD-9-CM Administrative Databases." *Journal of Clinical Epidemiology* 45 (6): 613–19.

Duan, N. 1983. "Smearing Estimate: A Nonparametric Retransformation Method." *Journal of the American Statistical Association* 78 (383): 605–10.

Edwards, N., D. Honemann, D. Burley, and M. Navarro. 1994. "Refinement of the Medicare Diagnosis–Related Groups to Incorporate a Measure of Severity." *Health Care Financing Review* 16 (2): 45–64.

Fetter, R. B., J. L. Freeman, H. Park, K. Schneider, J. Lichtenstein, and the Health Systems Management Group. 1989. *DRG Refinement with Diagnostic Specific Comorbidities and Complications: A Synthesis of Current Approaches to Patient Classification.* Final Report. New Haven, CT: Health Systems Management Groups, School of Organization and Management. Prepared for the Health Care Financing Administration under Cooperative Agreement Nos. 15-C-8930/1-01 and 17-C-98930/1-0251.

Fine, M. J., J. J. Orloff, D. Arisumi, G. D. Fang, V. C. Arena, B. H. Hanusa, V. L. Yu, D. E. Singer, and W. N. Kapoor. 1990. "Prognosis of Patients Hospitalized with Community-Acquired Pneumonia." *American Journal of Medicine* 88 (5N): 1N–8N.

Freeman, J. L., R. B. Fetter, H. Park, K. C. Schneider, J. L. Lichtenstein, W. A. Bauman, C. C. Duncan, J. S. Hughes, D. H. Freeman, Jr., and G. R. Palmer. 1991. "Refinement." In *DRGs: Their Design and Development*, edited by R. B. Fetter, Sr., D. A. Brand, and D. Gamache. Ann Arbor, MI: Health Administration Press.

Freeman, J. L., R. B. Fetter, H. Park, K. C. Schneider, J. L. Lichtenstein, J. S. Hughes, W. A. Bauman, C. C. Duncan, D. H. Freeman, Jr., and G. R. Palmer. 1995. "Diagnosis-Related Group Refinement with Diagnosis- and Procedure-Specific Comorbidities and Complications." *Medical Care* 33 (8): 806–27.

Gaul, G. M. 1990. "The Costs of Surgery Vary Greatly at Hospitals in Philadelphia Study Shows." *Philadelphia Inquirer* 13 December: 1-A.

Geraci, J. M., A. K. Rosen, A. S. Ash, K. J. McNiff, and M. A. Moskowitz. 1993. "Predicting the Occurrence of Adverse Events after Coronary Artery Bypass Surgery." *Annals of Internal Medicine* 118 (1): 18–24.

Gertman, P. M., and S. Lowenstein. 1984. "A Research Paradigm for Severity of Illness: Issues for the Diagnosis-Related Group System." *Health Care Financing Review* Suppl: 79–90.

Gonnella, J. S., M. C. Hornbrook, and D. Z. Louis. 1984. "Staging of Disease: A Case-Mix Measurement." *Journal of the American Medical Association* 251 (5): 637–44.

Gonnella, J. S., D. Z. Louis, C. Zeleznik, and B. J. Turner. 1990. "The Problem of Late Hospitalization: A Quality and Cost Issue." *Academic Medicine* 65 (5): 314–19.

Hawkes, J. 1989. "1st Consumer Guide to Hospitals Compares Costs, Results in Region." *Intelligencer Journal* 1 (June): 1.

Health Care Financing Administration, Office of Research and Demonstrations. 1987. *Report to Congress: DRG Refinement: Outliers, Severity of Illness, and Intensity of Care.* Publication No. (HCFA) 03254. Baltimore, MD: U.S. Department of Health and Human Services.

Health Systems Management Group, School of Organization and Management, Yale University. 1989. *DRG Refinement with Diagnostic Specific Comorbidities and Complications: A Synthesis of Current Approaches to Patient Classification.* New Haven, CT: Yale University. Prepared for the Health Care Financing Administration, under Cooperative Agreement #15-C-98930/1-01 and #17-C-98930/1-0251.

——. 1993. "Proposed Rules for the Medicare Program." *Federal Register:* 26 May.

Heuser, M. D., L. D. Case, and W. H. Ettinger. 1992. "Mortality in Intensive Care Patients with Respiratory Disease: Is Age Important?" *Archives of Internal Medicine* 152 (8): 1683–88.

Hoaglin, D. C., and B. Iglewicz. 1987. "Fine-Tuning Some Resistant Rules for Outlier Labeling." *Journal of the American Statistical Association.* 82 (December): 1147–49.

Horn, S. D., P. D. Sharkey, J. M. Buckle, J. E. Backofen, R. F. Averill, and R. A. Horn. 1991. "The Relationship Between Severity of Illness and Hospital Length of Stay and Mortality." *Medical Care* 29 (4): 305–17.

Hornbrook, M. C. 1985. "Techniques for Assessing Hospital Case Mix." *Annual Review of Public Health* 6: 295–324.

Iezzoni, L. I. 1994. *Risk Adjustment for Measuring Health Care Outcomes.* Ann Arbor, MI: Health Administration Press.

Iezzoni, L. I., A. S. Ash, J. L. Cobb, and M. A. Moskowitz. 1988. "Admission MedisGroups Score and the Cost of Hospitalizations." *Medical Care* 26 (11): 1069–80.

Iezzoni, L. I., A. S. Ash, G. Coffman, and M. A. Moskowitz. 1991a. "Admission and Mid-Stay MedisGroups Scores as Predictors of Death Within 30 Days of Hospital Admission." *American Journal of Public Health* 81 (1): 74–78.

——. 1991b. "Admission and Mid-Stay MedisGroups Scores as Predictors of Hospitalization Charges." *Medical Care* 29 (3): 210–20.

——. 1992. "Predicting In-Hospital Mortality: A Comparison of Severity Measurement Approaches." *Medical Care* 30 (4): 347–59.

Iezzoni, L. I., and L. G. Greenberg. 1994. "Widespread Assessment of Risk-Adjusted Outcomes: Lessons from Local Initiatives." *Joint Commission Journal of Quality Improvement* 20 (2): 305–16.

Iezzoni, L. I., E. K. Hotchkin, A. S. Ash, M. Shwartz, and Y. Mackiernan. 1993. "MedisGroups® Databases: The Impact of Data Collection Guidelines on Predicting In-Hospital Mortality." *Medical Care* 31 (3): 277–83.

Iezzoni, L. I., and M. A. Moskowitz. 1986. "A Clinical Assessment of MedisGroups." *Journal of the American Medical Association* 260 (21): 3159–63.

Iezzoni, L. I., M. Shwartz, A. S. Ash, J. S. Hughes, J. Daley, Y. Mackiernan, and D. Stone. 1995. *Evaluating Severity Adjustors for Patient Outcome Studies. Final Report.* Boston: Beth Israel Hospital. Prepared for the Agency for Health Care Policy and Research under Grant No. R01-HS06742.

Iezzoni, L. I., M. Shwartz, and J. Restuccia. 1991. "The Role of Severity Information in Health Policy Debates: A Survey of State and Regional Concerns." *Inquiry* 28 (2): 117–28.

Jencks, S. F., and A. Dobson. 1987. "Refining Case-Mix Adjustment: The Research Evidence." *The New England Journal of Medicine* 317 (11): 679–86.

Knaus, W. A., E. A. Draper, D. P. Wagner, and J. E. Zimmerman. 1985. "APACHE II: A Severity of Disease Classification System." *Critical Care Medicine* 13 (10): 818–29.

——. 1986. "An Evaluation of Outcome from Intensive Care in Major Medical Centers." *Annals of Internal Medicine* 104 (3): 410–18.

Knaus, W. A., D. P. Wagner, E. A. Draper, J. E. Zimmerman, M. Bergner, P. G. Bastos, C. A. Sirio, D. J. Murphy, T. Lotring, A. Damiano, and F. E. Harrell. 1991. "The APACHE III Prognostic System: Risk Prediction of Hospital Mortality for Critically Ill Hospitalized Adults." *Chest* 100 (6): 1619–36.

Knaus, W. A., D. P. Wagner, and J. Lynn. 1991. "Short-Term Mortality Predictions for Critically Ill Hospitalized Adults: Science and Ethics." *Science* 254 (5030): 389–94.

Knaus, W. A., D. P. Wagner, J. E. Zimmerman, and E. A. Draper. 1993. "Variations in Mortality and Length of Stay in Intensive Care Units." *Annals of Internal Medicine* 118 (10): 753–61.

Landis, J. R., and G. G. Koch. 1977. "The Measurement of Observer Agreement for Categorical Data." *Biometrics* 33 (1): 159–74.

MacKenzie, T. A., A. R. Willan, A. Greenaway-Coates, J. Comis, J. Fielding, J. Gerlach, A. Mahon, and R. Doran. 1991. *Patient Classification Systems: An Evaluation of the State of the Art, Volume I.* Kingston, Ontario: Case Mix Research, Queen's University.

Markson, L. E., D. B. Nash, D. Z. Louis, and J. S. Gonnella. 1991. "Clinical Outcomes Management and Disease Staging." *Evaluation and the Health Professions* 14 (2): 201–27.

McGuire, T. E. 1991. "An Evaluation of Diagnosis-Related Group Severity and Complexity Refinement." *Health Care Financing Review* 12 (4): 49–60.

Mendenhall, S. 1984. "DRGs Must Be Changed to Take Patient's Illness Severity into Account." *Modern Healthcare* 14: 86.

Naessens, J. M., C. L. Leibson, I. Krishan, and D. J. Ballard. 1992. "Contribution of a Measure of Disease Complexity (COMPLEX) to Prediction of Outcome and Charges Among Hospitalized Patients." *Mayo Clinic Proceedings* 67 (12): 1140–49.

Roos, L. L., S. M. Sharp, M. M. Cohen, and A. Wadja. 1989. "Risk Adjustment in Claims Based Research: The Search for Efficient Approaches." *Journal of Clinical Epidemiology* 42 (12): 1193–1206.

Schweiker, R. S., Secretary of Health and Human Services. 1982. *Report to Congress.* Washington, DC: Government Printing Office.

Silber, J. H., S. V. Williams, H. Krakauer, and J. S. Schwartz. 1992. "Hospital and Patient Characteristics Associated with Death after Surgery: A Study of Adverse Occurrence and Failure to Rescue." *Medical Care* 30 (7): 615–29.

Smith, D. W., M. Pine, R. C. Bailey, B. Jones, A. Brewster, and H. Krakauer. 1991. "Using Clinical Variables to Estimate the Risk of Patient Mortality." *Medical Care* 29 (11): 1108–29.

Smits, H. L., R. B. Fetter, and L. F. McMahon, Jr. 1984. "Variation in Resource Use Within Diagnosis-Related Groups: The Severity Issue." *Health Care Financing Review* Suppl: 71–78.

Steen, P. M., A. C. Brewster, R. C. Bradbury, E. Estabrook, and J. A. Young. 1993. "Predicted Probabilities of Hospital Death as a Measure of Admission Severity of Illness." *Inquiry* 30 (2): 128–41.

Thomas, J. W., and M. L. Ashcraft. 1991. "Measuring Severity of Illness: Six Severity Systems and Their Ability to Explain Cost Variations." *Inquiry* 28 (1): 39–55.

Thomas, J. W., M. L. F. Ashcraft, and J. E. Zimmerman. 1986. An Evaluation of Alternative Severity of Illness Measures for Use by University Hospitals. Technical Report, Vol. II. Department of Health Services Management and Policy, School of Public Health, The University of Michigan, Ann Arbor, MI.

U.S. Department of Health, Education and Welfare, National Committee on Vital and Health Statistics. 1980. *Uniform Hospital Discharge Data Minimum Data Set.* DHEW Publication No. (PHS) 80–1157. Hyattsville, MD: USDHEW.

Vladeck, B. C. 1984. "Medicare Hospital Payment by Diagnosis-Related Groups." *Annals of Internal Medicine* 100 (4): 576–91.

White House Domestic Policy Council. 1993. *Health Security: The President's Report to the American People.* Washington, DC: The Council.

Young, W. W. 1984. "Incorporating Severity of Illness and Comorbidity in Case-Mix Measurement." *Health Care Financing Review* Suppl: 23–31.

Young, W. W., S. Kohler, and J. Kowalski. 1994. "PMC Patient Severity Scale: Derivation and Validation." *Health Services Research* 29 (3): 367–90.

Young, W. W., R. B. Swinkola, and D. M. Zorn. 1982. "The Measurement of Hospital Case Mix." *Medical Care* 20 (5): 501–12.